

RESEARCH ARTICLE

Long-Range Correlations in Sentence Series from *A Story of the Stone*

Tianguang Yang¹, Changgui Gu², Huijie Yang^{2*}

1 Department of Statistics, School of Mathematical Sciences, Nankai University, Tianjin 300071, P. R. China, **2** Department of Systems Science, Business School, University of Shanghai for Science and Technology, Shanghai 200093, P. R. China

* hjyang@ustc.edu.cn



CrossMark
click for updates

OPEN ACCESS

Citation: Yang T, Gu C, Yang H (2016) Long-Range Correlations in Sentence Series from *A Story of the Stone*. PLoS ONE 11(9): e0162423. doi:10.1371/journal.pone.0162423

Editor: Zhong-Ke Gao, Tianjin University, CHINA

Received: May 6, 2016

Accepted: August 21, 2016

Published: September 20, 2016

Copyright: © 2016 Yang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available from the Website, <http://www.hongloumings.cn/>, and are in [S1 File](#).

Funding: This work is supported by the National Science Foundation of China under the Grant No. 10975099 to Professor Huijie Yang; National Natural Science Foundation of China 11505114 to Prof. Changgui Gu; the Shanghai Municipal Education Commission 13YZ072 to Professor Huijie Yang; the Shanghai Municipal Education Commission D-USST02 to Professor Huijie Yang; and the Shanghai Municipal Education Commission QD2015016 to Prof. Changgui Gu.

Abstract

A sentence is the natural unit of language. Patterns embedded in series of sentences can be used to model the formation and evolution of languages, and to solve practical problems such as evaluating linguistic ability. In this paper, we apply de-trended fluctuation analysis to detect long-range correlations embedded in sentence series from *A Story of the Stone*, one of the greatest masterpieces of Chinese literature. We identified a weak long-range correlation, with a Hurst exponent of 0.575 ± 0.002 up to a scale of 10^4 . We used the structural stability to confirm the behavior of the long-range correlation, and found that different parts of the series had almost identical Hurst exponents. We found that noisy records can lead to false results and conclusions, even if the noise covers a limited proportion of the total records (e.g., less than 1%). Thus, the structural stability test is an essential procedure for confirming the existence of long-range correlations, which has been widely neglected in previous studies. Furthermore, a combination of de-trended fluctuation analysis and diffusion entropy analysis demonstrated that the sentence series was generated by a fractional Brownian motion.

Introduction

Language is what makes us human [1, 2]. Over more than half a century, researchers have found rich patterns embedded in texts. For example, if we rank words in descending order of their frequencies, the frequency generally decreases according to a power-law of the rank, called Zipf's law [3]. Additionally, there are long-range correlations in the use of words [4–7]; words reoccur according to a stretched exponential distribution [8, 9]; and network based approaches can expose interesting characteristics of the relationships between words (e.g., small-world properties) [10–12]. These findings can shed light on the dynamical mechanisms behind the formation and evolution of languages. They also provide quantitative measures of language characteristics that can be used in various practical problems, such as distinguishing different authors and monitoring the development of linguistic abilities.

Relevant literature has mainly focused on words, but sentences are the natural unit of a text. Sentences supply a context to restrict a word to an exact meaning, and are organized logically

Competing Interests: The authors have declared that no competing interests exist.

into sentence groups, paragraphs, chapters, and entire texts to express ideas at different levels. Rather than syntactical principles, the sequential order of sentences is a result of many interacting factors such as logic, fluency, rhythm, harmony, intonation, and the author's style. A very recent paper by Drozd et al. [13] investigated series of sentence lengths and found almost perfect long-range correlations in more than one hundred classical novels from around the world.

It is well known that Chinese language is significantly different to western languages. In Chinese language, an elementary meaning is usually expressed with a single character instead of a word, and sometimes several characters are integrated into a group to express a unique meaning. Several short sentences with different subjects can occur within a single long sentence, if there are logical relationships. This is strictly forbidden in English. At a macroscopic level, Chinese people also have unique habits in terms of how they think. Hence, detecting scaling behaviors embedded in Chinese sentence series is an interesting and nontrivial task, and thus is the focus of this work. The contributions of our work can be summarized as follows.

We used de-trended fluctuation analysis (DFA) [14] to detect long-range correlations in the text of *A Story of the Stone*, which is one of the four most popular classical Chinese novels. We define the length of a sentence as the number of characters contained in the sentence. There exists a weak long-range correlation up to a scale of 10^4 (the Hurst exponent is 0.575 ± 0.002). The correlation structure is stable, i.e., different parts of the series have almost the same Hurst exponent. It is widely believed that this novel was written by two authors, i.e., the first to the 80th chapters were written by Xueqin Cao, and the 81th to the 120th chapters were written by E Gao, although there are still some disagreements [15]. However, we found no distinguishable difference in the long-range correlations of these two parts.

The 30th chapter of the novel uses the traditional Chinese punctuation symbol that represents a full stop (a small open circle), whereas the other chapters use the English full stop. This noise was used to test the impact of polluted data on long-range correlations detected using DFA. We found that introducing a small number of bad records (less than 1% of the total records) may result in incorrect conclusions. Therefore, to obtain reliable long-range correlation results we must test the structural stability of a series. To our knowledge, this procedure has been widely omitted in DFA calculations.

Furthermore, we combined DFA and diffusion entropy analysis (DEA) to find that the sentence series followed a fractional Brownian motion.

Materials and Methods

Data

A Story of the Stone is a masterpiece of Chinese vernacular literature and one of the four greatest Chinese classical novels [15]. It is remarkable not only for its huge cast of characters and psychological scope, but also for its precise and detailed observation of life and social structures in the 18th century China. It was written over approximately 10 years from 1749 to 1759. However, it was published anonymously until the 20th century, so there is some debate as to the author contributions. Currently, the 1st to 80th chapters are attributed to Xueqin Cao, and the remaining chapters are attributed to E Gao.

The text can be downloaded from the public repository *FigShare* with the accession number doi:[10.6084/m9.figshare.3759300](https://doi.org/10.6084/m9.figshare.3759300) (see also [S1 File](#) in the Supporting Information). From the text, we identified the positions of full stops (“.”), question marks (“?”), exclamation marks (“!”) and suspension points (“. . . .”). Then, we calculated the increments from the successive positions to form the sentence length series, each element of which is regarded as the length of the corresponding sentence. The sentence length series for the whole text contains 34,759

elements. The two parts attributed to Xueqin Cao and E Gao, hereafter called X-part and E-part, are 23,504 and 11,255 elements long, respectively.

Long-range correlations

Consider the sentence length series, $X = \{x_1, x_2, \dots, x_N\}$, where x_n is the length of the n th sentence, and N the total number of sentences. If X is stationary, a long-range power-law correlation implies that the autocorrelation function decays according to a power-law [16],

$$C(\tau) \equiv \frac{1}{N - \tau} \sum_{m=1}^{N-\tau} \frac{(x_m - \langle X \rangle)(x_{m+\tau} - \langle X \rangle)}{\sigma^2} \sim \tau^{-\alpha}, \tag{1}$$

where σ and $\langle X \rangle$ are the standard deviation and average of X , respectively.

The series X can be described using a random walk with displacements

$$y_n = \sum_{k=1}^n (x_k - \langle X \rangle), n = 1, 2, \dots, N, \tag{2}$$

i.e., the n th displacement is the summation of all the first to n th elements in X . This integrated series is called the profile series, denoted by Y . Using the profile series, we can also define a set of displacements in a specified duration τ , $\Delta y(\tau) = \{y_{1+\tau} - y_1, y_{2+\tau} - y_2, \dots, y_N - y_{N-\tau}\}$, and estimate the probability distribution function of Δy , denoted as $p(\Delta y)$. If X has the long-range correlation defined in Eq (1), the profile series will be scale-invariant (self-similar) [17], namely,

$$p[\Delta y(\tau)] \sim \frac{1}{\tau^\delta} F\left[\frac{\Delta y(\tau)}{\tau^\delta}\right], \tag{3}$$

where δ is the scaling exponent and $F(\cdot)$ a function. There exists a simple relation between δ and α , which is $\delta = 1 - \frac{\alpha}{2}$ [18, 19]. Hence, the scaling exponent δ can be used to measure long-range correlation behavior in the initial series, X . However, the sentence series (X) is generally non-stationary, so special procedures are required to evaluate the scaling exponent, δ .

De-trended fluctuation analysis. DFA is a widely used method for detecting long-range correlations embedded in non-stationary time series [14]. We briefly describe it, so that this paper is as self-contained as possible.

First, we extract all possible segments with a predefined length w (window size) from the profile series Y . That is,

$$Z_n = \{y_n, y_{n+1}, \dots, y_{n+w-1}\}, n = 1, 2, \dots, N - w + 1. \tag{4}$$

Second, we fit each segment with a q order polynomial function, $f(i, n) = \sum_{b=0}^q a_b(n) i^b$, where $a_b(n)$, $b = 0, 1, \dots, q$ are the fitting parameters for the n th segment, and $i = 1, 2, \dots, w$. The fitting curves are taken as estimates of trends in the corresponding segments. Subtracting the trends from the profile segments results in a set of segments of a stationary series,

$$S_n = \{y_n - f(1, n), y_{n+1} - f(2, n), \dots, y_{n+w-1} - f(w, n)\}, \tag{5}$$

$$n = 1, 2, \dots, N - w + 1.$$

If there are long-range correlations, the standard deviation, DFA_q , will obey a power-law,

$$DFA_q(w) \equiv \sqrt{\frac{\sum_{n=1}^{N-w+1} \sum_{k=1}^w [y_{n+k-1} - f(k, n)]^2}{w(N-w+1)}} \sim w^H. \tag{6}$$

For completely random, persistent, and anti-persistent series, we will have $H = 0.5$, $H > 0.5$, and $H < 0.5$, respectively.

In our calculations, the order of the polynomial function was $q = 2$. Because we are interested in scaling behaviors, the scale ranges, for instance, of 10^0-10^1 , 10^1-10^2 , 10^2-10^3 , and 10^3-10^4 should have the same contributions when estimating the Hurst exponent. The window sizes were $w = [1.2^m], m = 1, 2, \dots$, and $q + 3 \leq w \leq \frac{N}{3}$, where $[\cdot]$ is the integer part of a real number. If we plot $\log[F_q(w)]$ versus $\log(w)$, the points are distributed at almost identical intervals along the $\log(w)$ axis. Accordingly, in the least square estimate of the Hurst exponent, all the scales contain almost the same number of points and consequently have identical contributions.

Diffusion Entropy Analysis

Alternatively, the scaling exponent δ can be evaluated using diffusion entropy analysis (DEA) [20–22]. Starting from Eq (3), a straightforward computation leads to a simple relation of Shannon entropy versus the scaling exponent. That is,

$$E(\tau) \equiv - \int_{-\infty}^{+\infty} p[\Delta y(\tau)] \ln p[\Delta y(\tau)] d[\Delta y(\tau)] \sim A + \delta \ln \tau. \tag{7}$$

Hence, the slope of this relation is an unbiased estimate of δ if the sentence length series is stationary, which is generally not the case in reality.

We used the central moving average method [23–26] to extract the trend in the displacement series, $\Delta y(\tau)$. Let a window with size τ slide along this series. The average value of the covered displacements is regarded as the trend of the central displacement, namely, the $[\frac{\tau+1}{2}]$ th displacement, where $[\cdot]$ is the integer part of a real number. Subtracting the trend from the displacement series results in a de-trended displacement series with length $N - 2\tau + 1$.

We divided the interval so that the elements of the de-trended displacement series were distributed into $R(\tau)$ bins, and calculated the number of de-trended displacements occurring in every bin, denoted as $r(\tau, j), j = 1, 2, \dots, R(\tau)$. Then, the probability distribution can be approximated using

$$p(\tau, j) \approx \frac{r(\tau, j)}{N - 2\tau + 1}, j = 1, 2, \dots, R(\tau). \tag{8}$$

The Shannon entropy is then estimated using

$$DE(\tau) = - \sum_{j=1}^{R(\tau)} \frac{r(\tau, j)}{N - 2\tau + 1} \ln \left[\frac{r(\tau, j)}{N - 2\tau + 1} \right]. \tag{9}$$

If $DE(\tau)$ obeys the relation in Eq (7), the sentence length series is scale-invariant. This method is called diffusion entropy analysis (DEA).

Joint use of DFA and DEA

DEA can produce unbiased evaluation of δ for any kind of process, whereas the Hurst exponent (H) obtained by DFA is dynamical process dependent (this is why we use H instead of δ to represent the calculated scaling exponent in the DFA method) [22]. For a fractional Brownian motion, the estimated value of H is unbiased, namely, $H = \delta$. For a Levy walk process, we have a quantitative relation $\delta_{\text{levy}} = \frac{1}{3-2H} \leq H$. For a Levy flight process, the variance diverges and the DFA fails to qualitatively detect the scaling behavior. Hence, a combination of DEA and DFA (i.e., the relationship between H and δ) can provide information on the dynamic mechanism [22, 27–31].

Results

In the traditional Chinese punctuation system, a full stop is represented by a small open circle. However, with the increased computer usage, the English language full stop (“.”) has found its way into the Chinese language punctuation system. As a result, the two symbols have become acceptable. In the original downloaded text, the small open circle was used in the 30th chapter, while the rest of the chapters were punctuated with the English full stop (“.”) to signify the end of a sentence. We therefore recognize the symbol “.” as a full stop, and the small open circle as an independent Chinese character. This way, noise was introduced in the sentence series leading to a wrong number of identified sentences in the 30th chapter, and contaminating about 1% of the sentence lengths in the whole novel. Fig 1(a) presents the sentence length series. Most of the sentences are short, while some sentences are tens to hundreds of characters long. As shown in the inset panel of (a), some long sentences are clustered at around the 7850–7950th elements. This clustering comes from polluted records in the 30th chapter.

The sentence length obeys a right-skewed distribution as shown by the black solid circles in Fig 1(b). This distribution is log-normal (see the red solid circles in Fig 1(c)), i.e., the logarithm of sentence length is normally distributed. The characteristic length is 17. This distribution can

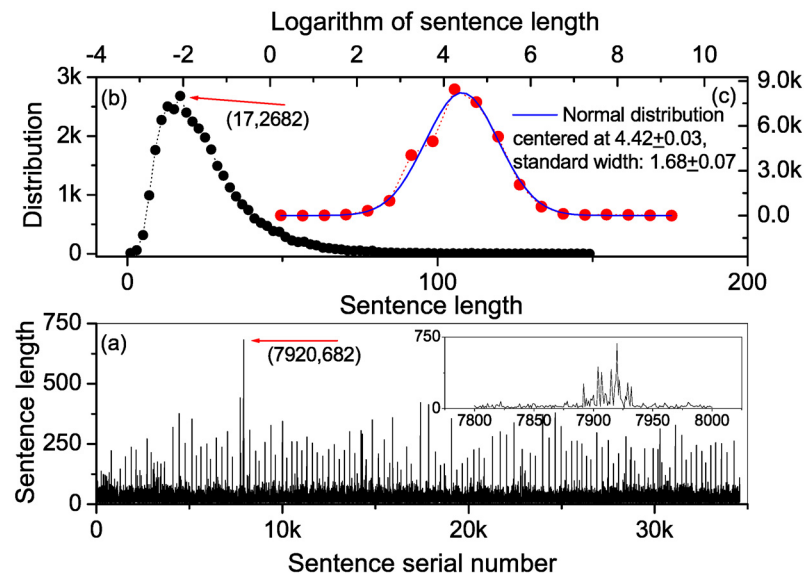


Fig 1. Statistics of sentence lengths. (a) Sentence length series. *Inset:* some sentences with large lengths are clustered together. This cluster is due to the traditional Chinese full stop symbol (small open circle) that was used in the 30th chapter, which produced noisy data. (b) Sentence length follows a right-skewed distribution, which is the log-normal distribution in (c).

doi:10.1371/journal.pone.0162423.g001

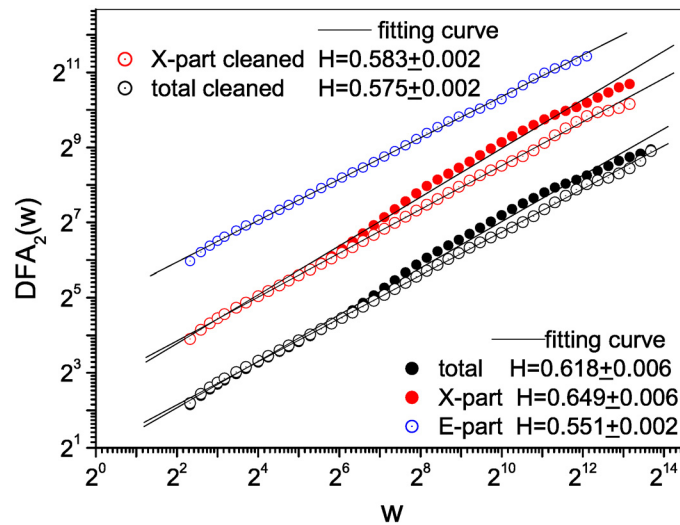


Fig 2. Long-range correlations in the noisy and cleaned series. Noisy records can result in incorrect estimates of the Hurst exponents (solid circles) and consequently false conclusions, even if they only cover a limited proportion of the total series. The effect of noise can be removed using a cleansing procedure (open circles). The X-and E-part of the text are the first to 80th chapters, and the 81th to the 120th chapters, which are currently attributed to Xueqin Cao and E Gao, respectively.

doi:10.1371/journal.pone.0162423.g002

be understood in terms of stochastic models of language [32, 33]. Consider a sentence with length L , and for simplicity assume that every character appears in a certain position of the sentence with identical probability, namely, $\frac{1}{L}$. The quantity of information contained in the sentence can be measured using Shannon entropy, $-\sum_{m=1}^L \frac{1}{L} \times \ln\left(\frac{1}{L}\right) = \ln L$. The log-normal distribution in Fig 1(c) tells us that the information contained in the sentences is distributed according to a Gaussian function.

Using the noisy series (extracted from the original text), we calculated DFA_2 for the total series, the series for the X-part (Chapters01–80), and the series for the E-part (Chapters81–120) as shown in Fig 2. The curves for the entire text and the X-part decrease slightly at large scales (black and red solid circles), but can be regarded as roughly obeying the power-law with Hurst exponents of $H = 0.618 \pm 0.006$ and $H = 0.649 \pm 0.006$, respectively. The E-part follows a perfect power-law relationship with a Hurst exponent of $H = 0.55 \pm 0.002$.

This result seems to be perfect, because we can conclude that the whole text, the X-part, and the E-part all have long-range correlations over a wide range (10^1 – 10^4). Furthermore, the X-part has a significantly high Hurst exponent compared with that of the E-part (the difference was $\Delta H = 0.649 - 0.55 \approx 0.1$). This appears to be because the two parts were written by different authors. Thus, it appears that we can distinguish between the language of the two authors.

But is this true? To answer this question, we used a cleansing procedure to filter out the effects of the noise. We replaced all the small open circles in the 30th chapter with the full stop “.” and re-extracted the sentence series, called cleaned series. As shown in Fig 2, the noise induced deformations were exactly corrected. The Hurst exponents were 0.575, 0.585, and 0.551 for the cleaned total, cleaned X-part, and the E-part (not polluted) respectively. They are almost identical, and the X-part and E-part are no longer distinguishable. So the polluted records may lead to exciting but incorrect conclusions. Hence, one must be careful when using the DFA method to detect long-range correlations, even when using sufficiently long series.

However, using the sentence length series in Fig 1(a), we cannot find significant evidence for the noise. Actually, when we investigate empirical series, there are typically some state

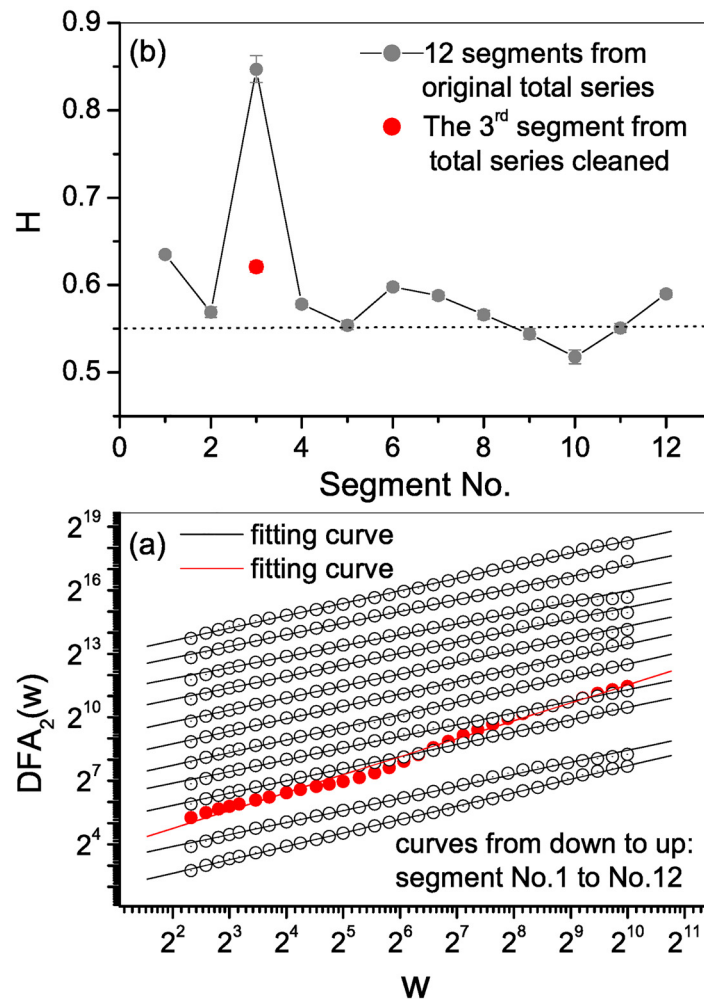


Fig 3. Structural stability test. The total polluted series was separated into 12 non-overlapping segments with a length of 2881. (a) All the curves obey almost perfect power-law relationships (gray open circles), except the 3rd segment which deviates significantly (red solid circles). (b) Hurst exponents for the 12 segments. The unreasonable large value of 0.87 for the 3rd noisy segment was corrected to 0.621 (red solid circle) after applying the cleaning procedure. There is a slightly decreasing trend.

doi:10.1371/journal.pone.0162423.g003

changes of the measured objects. For example, the physiological state of a volunteer may change occasionally in a long-term experiment, e.g., walking or sleeping. These occasional changes lead to abrupt changes in the records. But it is very difficult to find evidence of this sort of noise. Determining the correctness of the DFA calculations is non-trivial and cannot be solved using a single result for a series with a specific length. Herein, we will show that a structural stability test may help to confirm the existence of long-range correlations.

We separated the noisy series into 12 non-overlapping segments that were 2881 characters long. The third segment covers the noisy records. For every segment, we calculated the DFA_2 as shown in Fig 3(a). All the curves (gray open circles) almost exactly followed a power-law relationship, except for the 3rd segment (red solid circles), which significantly deviated from a straight line. A linear fit resulted in an unreasonably large estimated slope (0.87, see Fig 3(b)).

Using the cleaned series, there was a slightly decreasing trend in the scaling exponents. The unreasonable scaling exponent for the 3rd segment reduced to 0.621 (see Fig 3(b)—red solid circle), while the 8th to 12th segments exhibited comparatively smaller values. Since the novel

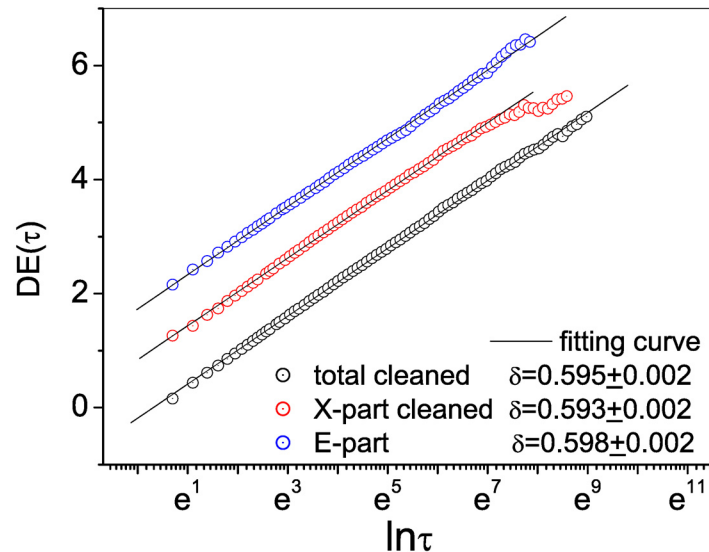


Fig 4. Diffusion entropy analysis of the cleaned total, cleaned X-part, and E-part series. Estimated values of the scaling exponent for the three series were almost identical.

doi:10.1371/journal.pone.0162423.g004

was written over a long period (approximately 10 years), it is difficult to attribute this finding to differences in the two authors' style or to a style change of a single author. Moreover, since most of the segments portray very similar scaling exponents, we conclude that they are structurally stable thus confirming the existence of long-range correlations in the whole series.

Additionally, the DEA method was used to detect the scaling behaviors embedded in the cleaned total, cleaned X-part, and E-part series. Fig 4 contains plots of diffusion entropy to $\ln \tau$ and the estimated values of the scaling-invariant exponent. For the total, X-part, and E-part series, the estimates were $(H, \delta) = (0.575, 0.595)$, $(0.583, 0.593)$ and $(0.551, 0.598)$, with corresponding values for $\frac{1}{3-2H}$ of 0.541, 0.545 and 0.527, respectively. Hence, we have $\frac{1}{3-2H} > H$ in all three cases. Because the error interval was $[-0.002, +0.002]$, the results support the conclusion of $\delta = H$. Consequently, the cleansed sentence series was generated by a fractional Brownian motion.

To ensure that the scaling behaviors were dependent on the sequential order of the sentences, we shuffled the sentence series and its segments and recalculated the scaling exponents. This resulted in $H = 0.5$ with confidence intervals less than $[-0.003, +0.003]$ (not shown).

Conclusion

In summary, non-trivial patterns in sentence length series can provide information on the structural behaviors of a text, which reflect characteristics of, for example, the author's style and logic. This information can be used to construct theoretical models of language formation and evolution, and also be helpful to solve practical problems such as evaluating linguistic ability and distinguishing different authors. Unlike most western languages, the Chinese language differs with respect to its organization at different levels. This work thus focused on a Chinese famous novel, *A Story of the Stone*, which is attributed to two different authors over a span of ten years.

Weak long-range correlations of the text were found using the DFA and DEA methods. The scaling behavior was perfect over a long scale covering approximately 10^1 – 10^4 . The scaling invariance was confirmed using the structural stability test. Although the Hurst exponent

slightly decreased with the sequential number of segments of the whole sentence series, we could not attribute this to the difference between two authors or to changes to a single author's style over ten years.

Results showed that when the DFA method is applied to noisy data, it can produce exciting but false results leading to incorrect conclusions. This was the case even if the noise affects a small proportion of the total series (e.g. 1%) and the series is sufficiently long (in this paper, $\sim 3 \times 10^4$). To confirm the existence of a long-range correlation, we should also apply the structural stability test to determine if different parts of the series have similar scale invariance. This procedure was rarely used in existing publications.

Furthermore, combining DFA and DEA we found that the sentence series of *A Story of the Stone* was produced by a fractional Brownian motion.

The structural stability test requires powerful new tools to detect scaling behaviors from short time series (see, for example, [34–41]). To reach further conclusions on the Chinese language, we will investigate more Chinese novels in the future.

Supporting Information

S1 File. S1_File.txt. Cleaned text of the *A Story of the Stone*. The data can also be downloaded from the public repository *FigShare*, uploaded with a title of *A Story of the Stone* and an accession number of doi:[10.6084/m9.figshare.3759300](https://doi.org/10.6084/m9.figshare.3759300).
(TXT)

S2 File. S2_File.m. Code for de-trended fluctuation analysis.
(M)

Acknowledgments

The authors thank Stephen Makau Mutua for his polishing the language.

Author Contributions

Conceptualization: HJY.

Data curation: TGY.

Formal analysis: TGY.

Funding acquisition: HJY CGG.

Investigation: TGY.

Methodology: TGY CGG.

Project administration: HJY.

Software: TGY.

Supervision: HJY.

Validation: CGG.

Visualization: TGY HJY.

Writing – original draft: HJY.

Writing – review & editing: HJY CGG.

References

1. Pinker S. The language instinct. NewYork, HarperCollins; 1994.
2. Bickerton D. Adam's tongue: how humans made language, how language made humans. NewYork, Hill and Wang; 2009.
3. Zipf GK. Human Behavior and the Principle of Least Effort. Addison-Wesley, Cambridge; 1949.
4. Montemurro MA, Pury PA. Long-range fractal correlations in literary corpora. *Fractals* 2002 Dec; 10:451–461. doi: [10.1142/S0218348X02001257](https://doi.org/10.1142/S0218348X02001257)
5. Altmann EG, Cristadoro G, Esposti MD. On the origin of long-range correlations in texts. *Proc. Natl. Acad. Sci.* 2012 Jul; 109:11582–11587. doi: [10.1073/pnas.1117723109](https://doi.org/10.1073/pnas.1117723109) PMID: [22753514](https://pubmed.ncbi.nlm.nih.gov/22753514/)
6. Ausloos M. Generalized hurst exponent and multifractal function of original and translated texts mapped into frequency and length time series. *Phys. Rev. E* 2012 Sept; 86:031108. doi: [10.1103/PhysRevE.86.031108](https://doi.org/10.1103/PhysRevE.86.031108)
7. Ausloos M. Measuring complexity with multifractals in texts: Translation effects. *Chaos, Solit. Fract.* 2012 Nov; 45(11):1349–1357. doi: [10.1016/j.chaos.2012.06.016](https://doi.org/10.1016/j.chaos.2012.06.016)
8. Laherrere J, Sornette D. Stretched exponential distributions in nature and economy: fat tails with characteristic scales. *Eur. Phys. J. B* 1998 April; 2(4):525–539. doi: [10.1007/s100510050276](https://doi.org/10.1007/s100510050276)
9. Altmann EG, Pierrehumbert JB, Motter AE. Beyond word frequency: bursts, lulls, and scaling in the temporal distributions of words. *Plos ONE* 2009 Nov; 4(11):e7678. doi: [10.1371/journal.pone.0007678](https://doi.org/10.1371/journal.pone.0007678) PMID: [19907645](https://pubmed.ncbi.nlm.nih.gov/19907645/)
10. Cancho RF-i, Sole RV. The small world of human language. *Proc. Roy. Soc. Lond. B* 2001 Nov; 268:2261–2265. doi: [10.1098/rspb.2001.1800](https://doi.org/10.1098/rspb.2001.1800)
11. Cong J, Liu H. Approaching human language with complex networks. *Phys. Life Rev.* 2014 Dec; 11:598618. And references there-in. doi: [10.1016/j.plrev.2014.04.004](https://doi.org/10.1016/j.plrev.2014.04.004)
12. Kulig A, Drozd S, Kwapien J, Oswiecimka P. Modeling the average shortest-path length in growth of word-adjacency networks. *Phys. Rev. E* 2015 Mar; 91:032810. doi: [10.1103/PhysRevE.91.032810](https://doi.org/10.1103/PhysRevE.91.032810)
13. Drozd S, Oswiecimka P, Kulig A, Wkwapien J, Bazarnik K, Grabska-Gradzinska I, et al. Quantifying origin and character of long-range correlations in narrative texts. *Information Sciences* 2016 Feb; 331:32–44. doi: [10.1016/j.ins.2015.10.023](https://doi.org/10.1016/j.ins.2015.10.023)
14. Peng CK, Buldyrev SV, Havlin S, Simons M, Stanley HE, Goldberger AL. Mosaic organization of DNA nucleotides. *Phys. Rev. E* 1994 Feb; 49:1685–1689. doi: [10.1103/PhysRevE.49.1685](https://doi.org/10.1103/PhysRevE.49.1685)
15. Zhou RC. Between Noble and Humble: Cao Xueqin and the Dream of the Red Chamber, edited by Ronald RG and Mark SF. New York: Peter Lang; 2009.
16. Stanley HE. Introduction to Phase Transitions and Critical Phenomena. Oxford University Press; 1971.
17. Mandelbrot BB. The Fractal Geometry of Nature. Freeman, San Francisco; 1982.
18. Buldyrev SV, Goldberger AL, Havlin S, Mantegna RN, Matsa ME, Peng CK, et al. Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis. *Phys. Rev. E* 1995 May; 51:5084–5091. doi: [10.1103/PhysRevE.51.5084](https://doi.org/10.1103/PhysRevE.51.5084)
19. Xiao Q, Pan X, Mutua S, Yang Y, Li XL, Yang HJ. Discrete scale-invariance in cross-correlations between time series. *Physica A* 2015 Mar; 421:161–170. doi: [10.1016/j.physa.2014.11.032](https://doi.org/10.1016/j.physa.2014.11.032)
20. Scafetta N, Hamilton P, Grigolini P. The thermodynamics of social processes: The teen birth phenomenon. *Fractals* 2001 Jun; 9(2):193–208. doi: [10.1142/S0218348X0100052X](https://doi.org/10.1142/S0218348X0100052X)
21. Grigolini P, Palatella L, Raffaelli G. Asymmetric anomalous diffusion: an efficient way to detect memory in time series. *Fractals* 2001 Dec; 9:439–449. doi: [10.1142/S0218348X01000865](https://doi.org/10.1142/S0218348X01000865)
22. Scafetta N, Grigolini P. Scaling detection in time series: Diffusion entropy analysis. *Phys. Rev. E* 2002 Sept; 66:036130. doi: [10.1103/PhysRevE.66.036130](https://doi.org/10.1103/PhysRevE.66.036130)
23. Alessio E, Carbone A, Castelli G, Frappietro V. Second-order moving average and scaling of stochastic time series. *Eur. Phys. J. B* 2002 May; 27:197–200.
24. Xu L, Ivanov PCh, Hu K, Carbone A, Stanley HE. Quantifying signals with power-law correlations: A comparative study of detrended fluctuation analysis and detrended moving average techniques. *Phys. Rev. E* 2005 May; 71:051101. doi: [10.1103/PhysRevE.71.051101](https://doi.org/10.1103/PhysRevE.71.051101)
25. Jiang ZQ, Zhou WX. Multifractal detrending moving-average cross-correlation analysis. *Phys. Rev. E* 2011 Jul; 84:016106. doi: [10.1103/PhysRevE.84.016106](https://doi.org/10.1103/PhysRevE.84.016106)
26. Gao ZK, Fang PC, Ding MS, Jin ND. Multivariate weighted complex network analysis for characterizing nonlinear dynamic behavior in two-phase flow. *Exp. Therm. & Flu. Sci.* 2015 Jan; 60:157–164. doi: [10.1016/j.exptthermflusci.2014.09.008](https://doi.org/10.1016/j.exptthermflusci.2014.09.008)

27. Scafetta N, West BJ. Multiscaling Comparative Analysis of Time Series and a Discussion on Earthquake Conversations in California. *Phys. Rev. Lett.* 2004 April; 92:138501. doi: [10.1103/PhysRevLett.92.138501](https://doi.org/10.1103/PhysRevLett.92.138501) PMID: [15089646](https://pubmed.ncbi.nlm.nih.gov/15089646/)
28. Scafetta N, West BJ. Solar flare intermittency and the Earth's temperature anomalies. *Phys. Rev. Lett.* 2003 Jun; 90:248701. doi: [10.1103/PhysRevLett.90.248701](https://doi.org/10.1103/PhysRevLett.90.248701) PMID: [12857233](https://pubmed.ncbi.nlm.nih.gov/12857233/)
29. Yang HJ, Zhao FC, Qi LY, Hu BL. Temporal series analysis approach to spectra of complex networks. *Phys. Rev. E* 2004 Jun; 69:066104. doi: [10.1103/PhysRevE.69.066104](https://doi.org/10.1103/PhysRevE.69.066104)
30. Scafetta N. Diffusion Entropy Analysis of Time Series: Theory, concepts, applications and computer codes for studying fractal noises and Levy walk signals. VDM Verlag Dr. Miller; Jul 2010.
31. Gao ZK, Yang YX, Zhai LS, Ding MS, Jin ND. Characterizing slug to churn flow transition by using multivariate pseudo Wigner distribution and multivariate multiscale entropy. *Chem. Engi. J.* 2016 May; 291:74–81. doi: [10.1016/j.cej.2016.01.039](https://doi.org/10.1016/j.cej.2016.01.039)
32. Russell DG, and Quentin DA. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 2003 Nov; 426:435–439. doi: [10.1038/nature02029](https://doi.org/10.1038/nature02029)
33. Peter F, and Alfred T. Toward a phylogenetic chronology of ancient Gaulish, Celtic, and Indo-European. *Proc. Natl. Acad. Sci. USA* 2003 Jul; 100:9079–9084. doi: [10.1073/pnas.1331158100](https://doi.org/10.1073/pnas.1331158100)
34. Qi JC, and Yang HJ. Hurst exponents for short time series. *Phys. Rev. E* 2011 Dec; 84:066114. doi: [10.1103/PhysRevE.84.066114](https://doi.org/10.1103/PhysRevE.84.066114)
35. Zhang WQ, Qiu L, Xiao Q, Yang HJ, Zhang QJ, and Wang JY. Evaluation of scale invariance in physiological signals by means of balanced estimation of diffusion entropy. *Phys. Rev. E* 2012 Nov; 86:056107. doi: [10.1103/PhysRevE.86.056107](https://doi.org/10.1103/PhysRevE.86.056107)
36. Pan X, Hou L, Stephen M, Yang HJ, and Zhu CP. Evaluation of scaling invariance embedded in short time series. *Plos ONE* 2014 Dec; 9(12):e116128. doi: [10.1371/journal.pone.0116128](https://doi.org/10.1371/journal.pone.0116128) PMID: [25549356](https://pubmed.ncbi.nlm.nih.gov/25549356/)
37. Pan X, Hou L, Stephen M, and Yang HJ. Long-term memories in online users' selection activities. *Phys. Lett. A* 2014 Jul; 378(35):2591–2596. doi: [10.1016/j.physleta.2014.07.012](https://doi.org/10.1016/j.physleta.2014.07.012)
38. Gao ZK, Jin ND. A directed weighted complex network for characterizing chaotic dynamics from time series. *Nonlinear Analysis: Real World Applications* 2012, April; 13:947–952. doi: [10.1016/j.nonrwa.2011.08.029](https://doi.org/10.1016/j.nonrwa.2011.08.029)
39. Gao ZK, Yang YX, Fang PC, Jin ND, Xia CY, Hu LD. Multi-frequency complex network from time series for uncovering oil-water flow structure. *Sci. Rept.* 2015 Feb; 5:8222. doi: [10.1038/srep08222](https://doi.org/10.1038/srep08222)
40. Mutua S, Gu CG, Yang HJ. Visibility Graph Based Time Series Analysis. *Plos ONE* 2015 Nov; 10(11):e0143015. doi: [10.1371/journal.pone.0143015](https://doi.org/10.1371/journal.pone.0143015)
41. Mutua S, Gu CG, Yang HJ. Visibility Graphlet Approach to Chaotic Time Series. *Chaos* 2016 May; 26(5):053107. doi: [10.1063/1.4951681](https://doi.org/10.1063/1.4951681) PMID: [27249947](https://pubmed.ncbi.nlm.nih.gov/27249947/)